SACRO: Semi-Automated Checking of Research Outputs

Professor Jim Smith,

University of the West of England
SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
(Semi) Automating this bottleneck!

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Who are we? (alphabetically)
Who are we? (alphabetically)

Universities

- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE
Who are we? (alphabetically)

<table>
<thead>
<tr>
<th>Universities</th>
<th>Public Data Bodies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aberdeen</td>
<td>Health Data Research UK</td>
</tr>
<tr>
<td>Dundee</td>
<td>NHS Scotland</td>
</tr>
<tr>
<td>Durham</td>
<td>Public Health Scotland</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>Research Data Scotland</td>
</tr>
<tr>
<td>Oxford</td>
<td></td>
</tr>
<tr>
<td>UWE</td>
<td></td>
</tr>
</tbody>
</table>

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Who are we? (alphabetically)

### Universities
- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

### Public Data Bodies
- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

### TREs
- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)
Who are we? (alphabetically)

Universities
- Aberdeen
- Dundee
- Durham
- Edinburgh
- Oxford
- UWE

Public Data Bodies
- Health Data Research UK
- NHS Scotland
- Public Health Scotland
- Research Data Scotland

TREs
- DASH (Aberdeen/Grampian)
- DataLoch (Edinburgh)
- HIC (Dundee)
- eDRIS (Public Health Scot)
- OpenSafely (Oxford)

External steering group:
UK: Office for National Statistics, ESRC, DARE
Global: Eurostat, SDC-GESIS, ICPSR (US)

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation
The current situation
The current situation

Confidential data in TRE
The current situation

Confidential data in TRE

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation
The current situation

Confidential data in TRE

Analysis

Stata, R, Python

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation

Confidential data in TRE

Analysis

Stata, R, Python

File1

File2

File3

Filen

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation

Confidential data in TRE

Stata, R, Python

Analysis

File 1

request

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation

Confidential data in TRE

Analysis

Stata, R, Python

request

File1
File2
File3
File4
Filen

TRE staff

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
The current situation

TRE staff

request

decision

File1

File1

Filen

Stata,R,Python

Analysis

Confidential data in TRE

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell
SACRO in a nutshell
SACRO in a nutshell

Confidential data in TRE

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

Confidential data in TRE

TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

Confidential data in TRE
TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

Confidential data in Stata, R, Python

TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

Confidential data in Stata, R, Python

ACRO engine

Confidential data in TRE

TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

TRE staff request

Analysis

Stata, R, Python

ACRO engine

Confidential data in TRE

TRE risk appetite

ACRO report

File1
File1
File1
Filene
SACRO in a nutshell

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO in a nutshell

ACRO engine

Stata, R, Python

Analysis

Confidential data in TRE

TRE risk appetite

File1 File1 Fileen

TRE staff request decision

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Similar for Machine Learning Models

Except that we:
Similar for Machine Learning Models

Except that we:

- Run a range of “inference” attacks
Similar for Machine Learning Models

Except that we:

• Run a range of “inference” attacks”

• Aim to support more ‘user journeys’
Similar for Machine Learning Models

Except that we:

• Run a range of “inference” attacks
• Aim to support more ‘user journeys’
• Don’t have a set of ‘tried and trusted’ guidelines to work with
ACRO: analytic commands
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

- `acro.logit()`, `acro.logitr()`
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

- `acro.logit()`, `acro.logitr()`
- `acro.ols()`, `acro.olsr()`
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

- `acro.logit()` , `acro.logitr()`
- `acro.ols()` , `acro.olsr()`
- `acro.probit()` , `acro.probitr()`
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

- `acro.logit()` , `acro.logitr()`
- `acro.ols()` , `acro.olsr()`
- `acro.probit()` , `acro.probitr()`

R versions are prefixed by “acro_”
**ACRO: analytic commands**

Using the same api as pandas, adding checks for cell count and dominance

- `acro.crosstab()`
- `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

- `acro.logit() , acro.logitr()`
- `acro.ols() , acro.olsr()`
- `acro.probit() , acro.probitr()`

R versions are prefixed by "acro_"

- Also support R’s built in ‘**table**’ command
ACRO: analytic commands

Using the same api as pandas, adding checks for cell count and dominance

• `acro.crosstab()`
• `acro.pivot_table()`

Using the same api as statsmodels, adding checks for DoF:

• `acro.logit()` , `acro.logitr()`
• `acro.ols()` , `acro.olsr()`
• `acro.probit()` , `acro.probitr()`

R versions are prefixed by “acro_”

• Also support R’s built in ‘table’ command

• Stata versions prefixed by “acro ”

• Code currently captures and translates commands: `table`, `regress`, `probit`, `logit`
ACRO: user commands for session management
ACRO: user commands for session management
ACRO: user commands for session management

- acro = ACRO()
ACRO: user commands for session management

- acro = ACRO()
- acro.suppress = [True, False]
ACRO: user commands for session management

- `acro = ACRO()`
- `acro.suppress = [True, False]`
- `acro.print_outputs()`
ACRO: user commands for session management

- acro = ACRO()
- acro.suppress = [True, False]
- acro.print_outputs()
- acro.remove_output(key)
ACRO: user commands for session management

• acro = ACRO()
• acro.suppress= [True, False]
• acro.print_outputs()
• acro.remove_output(key)
• acro.rename_output(key, newname)
**ACRO: user commands for session management**

- `acro = ACRO()`
- `acro.suppress= [True, False]`
- `acro.print_outputs()`
- `acro.remove_output(key)`
- `acro.rename_output(key, newname)`
- `acro.add_exception(key, text)`
ACRO: user commands for session management

- acro = ACRO()
- acro.suppress= [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
ACRO: user commands for session management

- `acro = ACRO()`
- `acro.suppress = [True, False]`
- `acro.print_outputs()`
- `acro.remove_output(key)`
- `acro.rename_output(key, newname)`
- `acro.add_exception(key, text)`
- `acro.add_comment(key, text)`
- `acro.custom_output(filename, description)`
ACRO: user commands for session management

- acro = ACRO()
- acro.suppress = [True, False]
- acro.print_outputs()
- acro.remove_output(key)
- acro.rename_output(key, newname)
- acro.add_exception(key, text)
- acro.add_comment(key, text)
- acro.custom_output(filename, description)
- acro.finalise(output_directory, format)
Example: what happens if researcher requests a disclosive table

```r
: suppress = False

safe_table = acro.crosstab(df.recommend, df.parents)
print(safe_table)
```

```r
INFO:acro:get_summary(): fail; threshold: 4 cells may need suppressing;
INFO:acro:outcome_df:

<table>
<thead>
<tr>
<th>parents</th>
<th>great_pret</th>
<th>pretentious</th>
<th>usual</th>
</tr>
</thead>
<tbody>
<tr>
<td>recommend</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| not_recom | ok | ok | ok |
| priority   | ok | ok | ok |
| recommend  | threshold; | threshold; | threshold; |
| spec_prior | ok | ok | ok |
| very_recom | threshold; | ok | ok |

INFO:acro:records:add(): output_1

<table>
<thead>
<tr>
<th>parents</th>
<th>great_pret</th>
<th>pretentious</th>
<th>usual</th>
</tr>
</thead>
<tbody>
<tr>
<td>recommend</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not_recom</td>
<td>1440</td>
<td>1440</td>
<td>1440</td>
</tr>
<tr>
<td>priority</td>
<td>858</td>
<td>1484</td>
<td>1924</td>
</tr>
<tr>
<td>recommend</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>spec_prior</td>
<td>2022</td>
<td>1264</td>
<td>758</td>
</tr>
<tr>
<td>very_recom</td>
<td>0</td>
<td>132</td>
<td>196</td>
</tr>
</tbody>
</table>
```
SACRO viewer for TRE output checkers

SACRO Outputs Viewer

output_1
- crosstab table

output_3
- pivot_table table
- cts regression
- ols regression
- probit regression

output_6
- logit regression
- crosstab table
- custom

output_7


ACRO risk profile

output_1
- Created at: 21 August 2023 at 12:16
- Output type: crosstab table
- ACRO status: Fail (threshold: 4 cells may need suppressing)
- Comments:
  - Please let me have this data.
  - 6 cells were suppressed in this table
- Exception request: because I really want it
- Review: You cannot approve this output until you add a comment.
  - Approve
  - Reject

Review comments on output_1:

output_1.csv

<table>
<thead>
<tr>
<th>recommend</th>
<th>great_pret</th>
<th>pretentious</th>
<th>usual</th>
</tr>
</thead>
<tbody>
<tr>
<td>not_recom</td>
<td>1440</td>
<td>1440</td>
<td>1440</td>
</tr>
<tr>
<td>priority</td>
<td>858</td>
<td>1484</td>
<td>1924</td>
</tr>
<tr>
<td>recommend</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>spec_prior</td>
<td>2022</td>
<td>1264</td>
<td>758</td>
</tr>
<tr>
<td>very_recom</td>
<td>0</td>
<td>132</td>
<td>196</td>
</tr>
</tbody>
</table>

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO viewer for TRE output checkers

Outputs to review

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO viewer for TRE output checkers

Outputs to review

Output with problematic cells highlighted
SACRO viewer for TRE output checkers

Outputs to review:

Output with problematic cells highlighted:

Review TRE risk appetite:

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO viewer for TRE output checkers

Outputs to review

Type of output
Recommendation
Comments from user
Exception. request

view TRE risk appetite

Output with problematic cells highlighted

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
SACRO viewer for TRE output checkers

- Outputs to review
- Output with problematic cells highlighted
- View TRE risk appetite
- Type of output
- Recommendation
- Comments from user
- Exception request
- Comments needed to override recommendation
SACRO viewer for TRE output checkers

- Outputs to review
- Output with problematic cells highlighted
- SACRO Outputs Viewer
  - Type of output
  - Recommendation
  - Comments from user
  - Exception request
  - Comments needed to override recommendation
  - Record overall comments, create release package
  - View TRE risk appetite

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Get Involved

Get involved

Thanks to:
Get Involved

Get involved

ACRO ‘engine’:

- [https://github.com/AI-SDC/ACRO](https://github.com/AI-SDC/ACRO)

Thanks to:
Get Involved

ACRO ‘engine’:
- https://github.com/AI-SDC/ACRO

Viewer:
- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk
Get Involved

Get involved

ACRO ‘engine’:

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Get Involved

Get involved

ACRO ‘engine’:
- [https://github.com/AI-SDC/ACRO](https://github.com/AI-SDC/ACRO)

Viewer:
- [https://github.com/opensafely-core/sacro](https://github.com/opensafely-core/sacro)

Consensus statement: [sacro.contact@uwe.ac.uk](mailto:sacro.contact@uwe.ac.uk)

Anything else: [sacro.contact@uwe.ac.uk](mailto:sacro.contact@uwe.ac.uk)

Thanks to:
- All the project partners and TRE staff for all their feedback on the tools as they developed
Get Involved

Get involved

ACRO ‘engine’:
• https://github.com/AI-SDC/ACRO

Viewer:
• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Thanks to:
• All the project partners and TRE staff for all their feedback on the tools as they developed
• Members of the public for input to consensus statement

Anything else: sacro.contact@uwe.ac.uk
Get Involved

Get involved

ACRO ‘engine’:

• [https://github.com/AI-SDC/ACRO](https://github.com/AI-SDC/ACRO)

Viewer:

• [https://github.com/opensafely-core/sacro](https://github.com/opensafely-core/sacro)

Consensus statement: [sacro.contact@uwe.ac.uk](mailto:sacro.contact@uwe.ac.uk)

Anything else: [sacro.contact@uwe.ac.uk](mailto:sacro.contact@uwe.ac.uk)

Thanks to:

• All the project partners and TRE staff for all their feedback on the tools as they developed

• Members of the public for input to consensus statement

• External steering group
Get Involved

Get involved

ACRO ‘engine’:

- https://github.com/AI-SDC/ACRO

Viewer:

- https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:

- All the project partners and TRE staff for all their feedback on the tools as they developed
- Members of the public for input to consensus statement
- External steering group
- You for listening
Get Involved

Get involved

ACRO ‘engine’:
• https://github.com/AI-SDC/ACRO

Viewer:
• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:
• All the project partners and TRE staff for all their feedback on the tools as they developed
• Members of the public for input to consensus statement
• External steering group
• You for listening

SACRO: Professor Jim Smith, UNECE Expert meeting on Statistical Data Confidentiality 2023
Get Involved

Get involved

ACRO ‘engine’:
• https://github.com/AI-SDC/ACRO

Viewer:
• https://github.com/opensafely-core/sacro

Consensus statement: sacro.contact@uwe.ac.uk

Thanks to:
• All the project partners and TRE staff for all their feedback on the tools as they developed
• Members of the public for input to consensus statement
• External steering group
• You for listening

Anything else: sacro.contact@uwe.ac.uk
Get Involved

Get involved

ACRO ‘engine’:
• [https://github.com/AI-SDC/ACRO](https://github.com/AI-SDC/ACRO)

Viewer:
• [https://github.com/opensafely-core/sacro](https://github.com/opensafely-core/sacro)

Consensus statement: sacro.contact@uwe.ac.uk

Anything else: sacro.contact@uwe.ac.uk

Thanks to:
• All the project partners and TRE staff for all their feedback on the tools as they developed
• Members of the public for input to consensus statement
• External steering group
• You for listening

This work is funded by UK research and Innovation, [Grant Number MC_PC_23006], as part of Phase 1 of the DARE UK (Data and Analytics Research Environments UK) programme, delivered in partnership with Health Data Research UK (HDR UK) and Administrative Data Research UK (ADR UK)